RADC-TDR-64-101
Final Report

ANALYSIS OF POLYGRAPHIC DATA

TECHNICAL DOCUMENTARY REPORT NO. RADC-TDR-64-101
January 1965

QRC Branch
Rome Air Development Center
Research and Technology Division
Air Force Systems Command
Griffiss Air Force Base, New York

Project No. 5534, Task No. 553401

(Prepared under Contract No. AF 30(602)-2634 by Joseph F. Kubis, Fordham University, New York, N.Y.)

FOREWORD

This report was prepared for RADC under Contract No. AF30(602)-2634, by Dr. J. F. Kubis of the Department of Psychology of Fordham University.

Many people have directly or indirectly contributed to the completion of this difficult assignment. Specially helpful has been the untiring cooperation of Mr. Rudolph Eckhardt who was intimately involved in most of the details of this pro Sincere thanks also go to Mr. Louis Hsu, Mr. Ronald Salafia, and Mr. Robert Zenhausern for their unstinting cooperation throughout the project. Special tha are due Dr. Donald Sweeney for help in planning the sample and in facilitating rating procedure during the early phase of our work.

Most of our gratitude goes to Mr. Robert Byrne, Quick Reaction Capability Laboratory, RADC for his untiring efforts to obtain facilities for our work a. his understanding of the difficulties we were continually encountering. Equally appreciated is the deep understanding and continued encouragement of Dr. She MacLeod, Human Engineering Laboratory, RADC.

AIR FORCE, Cato Show Prtg. Co. 2/26/65 - 150 -

ABSTRACT

To evaluate the feasibility of adapting rapid processing techniques to the present "art" of polygraph interpretation, it was necessary to evaluate and study the various factors affecting the processes of interpretation of polygraphic data. This report covers two sections dealing with this problem.

SECTION I: Factors Affecting the Decision Process in Lie Detection

Two types of decision situations are characteristic of lie detection investigations: the dependent judgment case in which the examiner, after comparing all records, selects the guilty individual (and possibly accomplices) from among a group of suspects known to include the culprit(s): and, the independent judgment case, in which a decision of innocence or guilt is made independently for each suspect on the basis of his record alone. In the latter situation the suspects are usually apprehended one at a time and at irregular intervals.

Rater accuracy for each decision situation was evaluated by using 100 of 336 records obtained in the Simulated Theft Experiment (Kubis, 1962), a dependent judgment situation. These records were evaluated under independent judgment conditions by three new raters and by one rater who also served as an examiner in the Simulated Theft Experiment. It was anticipated that the opportunity of comparing the records of all suspects in the dependent judgment situation would result in greater accuracy than that attainable in the independent judgment situation.

The results indicate that neither accuracy of decisions nor confidence in them was diminished under independent judgment conditions. However, the one rater who served in both experimental situations showed less accuracy and less confidence in his decisions in the independent judgment situation. Furthermore, the more "serious" errors of misclassification were more numerous in the independent judgment situation. Greatest accuracy was achieved with the psychogalvanic index of deception, and this index tended to determine the direction of the final decision in the analysis of the total polygraph chart.

SECTION II: Accuracy of Measured and Rated Physiological Response Systems Used in Lie Detection

Records of 33 subjects from the Simulated Theft Experiment were selected for further analysis and measurement of the three physiological response systems. The characteristics of the psychogalvanic response selected for measurement were relative change in resistance (Height) and recovery time (Width). Amplitude and Frequency were the measured indices obtained from the respiratory tracings. Height, Width and Change were measured from the plethysmographic response. The accuracies of these indices, separately or in combination, were compared with the accuracies attained by the ratings of lie detector operators who evaluated the total response pattern of each physiological response in arriving at their ratings.

The measured characteristics of the physiological response systems were found to be as accurate as the ratings of the lie detector operators in discriminating between culprit, collaborator, and innocent suspect. Continued research should make it possible to objectify most of the lie detection indices with the aid of a computer.

PUBLICATION REVIEW

Publication of this technical documentary report does not constitute Air Force approval of the report's findings or conclusions. It is published only for the exchange and stimulation of ideas.

Approved: Robert D. Byrne.
ROBERT D. BYRNE
Project Engineer

Approved:
ROBERT J. QUINN, JR.
Colonel, USAF
Chief, Intel & Info Processing Div

FOR THE COMMANDER:
IRVING J. GABELMAN
Chief, Advanced Studies Group

TABLE OF CONTENTS

LIST OF TABLES

SECTION I

FACTORS AFFECTING THE DECISION PROCESS IN LIE DETECTION

Two types of decision situations confront the so-called lie detector expert. In the one, he is called upon to examine a relatively small and fixed group of suspects. His objective is to determine the culprit among them. He is assured that these are the only suspects who could be associated with the crime. As an example, bank losses very often can be confined to a small area and to a small group of employees who could have had access to a particular safe or vault. In identifying the culprit, the expert is influenced by and dependent upon the mutual comparison of the polygraph charts of all suspects. This will be called a Dependent Judgment situation.

By contrast, the second type of case involves the examination of a single suspect. If there are more, they are brought in at irregular intervals, usually one at a time. A decision is rendered after the examination of each suspect. Guilt or innocence is determined independently for each suspect on the basis of his records alone. Naturally enough, this will be called the Independent Judgment situation.

In a previous research (Kubis, 1962) several aspects of the decision process in lie detection were studied. A Simulated Theft provided a situation in which a Thief (T), a Lookout (L) or confederate, and an Innocent Suspect (I) were involved. An examiner, playing the role of a lie detector expert, tested the three members of the Simulated Theft group immediately after the theft was committed. He knew that one of the three individuals to be examined was a Thief, one a Lookout, and one an Innocent Suspect. It was his job to identify the role of each suspect. After he tested the group of three, the examiner rated the physiological reactions to those questions that were directly related to the theft. The instrument he used recorded respiratory changes (Resp), a plethysmographic pattern (Pleth), and the psychogalvanic reaction (PGR). The examiner studied the physiological responses and made a decision as to the role each suspect assumed in the experiment. In other words, he tried to identify the Thief, the Lookout, and the Innocent Suspect on the basis of all the chart recordings he had just obtained. Having made his decision, the examiner then indicated the degree of his confidence in them.

STATEMENT OF PROBLEM

In actual circumstances the lie detector expert is not usually confronted with a small group of suspects among whom the guilty one is certain to be found. Often he examines a single individual and is asked for his decision after the examination. Furthermore, there are groups of suspects brought in for examination that do not have a culprit among them. Fundamentally, the expert must be prepared to make a decision of guilt or innocence (more accurately, of lying or truth-telling) in single cases, without having the opportunity of comparing the records of several suspects.

An important question, however, needs to be answered. How does the accuracy of the lie detector expert compare (a) in cases where there is but one suspect, (b) in cases where there are several suspects, one of whom being definitely guilty? One would intuitively expect greater accuracy in the latter situation. In terms of the Simulated Theft Experiment the question becomes: Would a rater, making a decision on each polygraph record singly and independently of other records, be as accurate as the raters in the Simulated Theft Experiment. The latter worked with and compared the records of all three suspects before arriving at their decisions. The problem is one of determining the relative accuracies of lie detection decisions in the Independent Judgment and Dependent Judgment situations.

PROCEDURE

Since all records from the Simulated Theft Experiment were available, it was a simple task to recode them after eliminating any markings that would identify the suspect or the examiner who did the testing. In this form the records could be reassembled and presented singly to a rater for a decision as to the role the subject played in the experiment. The accuracy of such ratings could be compared with the accuracy already reported in the Simulated Theft Experiment.

Of the five examiners who conducted the tests in the Simulated Theft Experiment and who also served as raters only one, Rater E, remained. For the present experiment one graduate student, Rater H, was carefully trained to interpret the polygraph charts, to operate the polygraph, and to administer the lie detection test. Two other graduate students, Raters Y and Z, were trained only up to the

level of chart interpretation. As yet these two had no practical experience; they were not trained in the use of the polygraph; they had not served as examiners in a lie detection experiment. There were, then, four raters two of whom were at a lower level of experience, namely, the level of chart interpretation.

Of the 336 records in the Simulated Theft Experiment, 100 were selected for the present experiment. To compare how accurately the same person would rate a set of records under both Dependent Judgment and Independent Judgment conditions, all 100 records were those in which Rater E, either as examiner or as rater, had been involved in the Simulated Theft Experiment. At no time was he aware that any specific record was one he had rated before. All he knew was that 100 of 336 records from the old experiment were included in this decision task. Recognition of specific peculiarities or clues was not highly probable since he had not seen his old records in over a year. Neither was it likely that he had lost his skills. Since the completion of the Simulated Theft Experiment he had been involved in numerous rating and training tasks related to lie detection.

To continue with the description of the 100 records. These included 10 complete groups (each with a Thief, a Lookout, and an Innocent Suspect) for which Rater E had served as examiner, i. e., the person who tested the subjects by means of the polygraph. An additional 23 complete groups (69 subjects tested by other examiners) were included because Rater E had rated them. One subject was randomly selected to round off the number to 100. The 100 records were placed into 10 large folders, each of which served to contain a convenient unit of work. No folder contained more than seven subjects with the same role. Each folder represented all three roles.

A random assignment of records to each folder was stressed in the directions to the raters. The purpose was to prevent an expectation of equal division of roles among the 100 records. At no time were the raters aware of the fact that entire groups (Thief, Lookout, Innocent Suspect) were selected from the Simulated Theft data.

The four raters for this experiment, Raters E, H, Y, and Z, were instructed to work independently and to evaluate one record at a time. The first task was to rate the respiratory response alone. This was accomplished by blocking out the

plethysmographic and psychogalvanic tracings. After completing his ratings on a particular record, the rater had to decide whether the person was a Thief, a Lookout, or an Innocent Suspect. He continued in this fashion until all 100 records were rated. The entire process was repeated for the plethysmographic tracings; and again, for the psychogalvanic response. Finally, the decision procedure was completed with the total record (respiratory patterns, plethysmographic tracings, and psychogalvanic reactions) exposed for analysis and available for interpretation. In all, each rater made 400 decisions, four for each record.

Only one decision was required in the original Simulated Theft Experiment, and this was based on an overall evaluation of the polygraph chart which included respiratory, plethysmographic and psychogalvanic tracings. In the present experiment, four independent decisions were required, one each for the separate physiological indices and a final one on the overall aspects of the total record. Consequently, only the overall evaluations in both experiments could be compared to assess the relative accuracies of decisions under Independent Judgment and Dependent Judgment conditions.

The comparative analyses discussed in the next section are based on 99 subjects, since the Dependent Judgment decisions can only come from an entire group involving a Thief, a Lookout, and an Innocent Suspect. The extra subject, included originally to fill out a folder of 10 subjects, was dropped from the analysis.

RESULTS

The purpose of this experiment was to compare diagnostic accuracy of judges under two decision conditions. In the Dependent Judgment situation, typified by the Simulated Theft Experiment, judges had before them records of a complete group consisting of a Thief, a Lookout, and an Innocent Suspect. After an evaluation of each record and a comparison of all, they had to identify which record belonged to the Thief, which to the Lookout, and which to the Innocent Suspect. Under Independent Judgment conditions raters examined and decided the status of one record at a time. The order in which the records were examined was random. These raters, then, seemed to operate with less information than that available to the raters in the Dependent Judgment situation.

An allied objective was to evaluate several factors that might possibly differentiate between these two types of decision situations. There was the matter of confidence in one's decisions, the nature of the errors made, and the factor of experience.

ACCURACY

Rater accuracy under Dependent and Independent Judgment conditions is presented in Table 1. Decisions were based on the total polygraph chart including all three indices -- respiratory, plethysmographic, and psychogalvanic. The accuracy scores were obtained from the records of the same sample of 99 subjects, as they were evaluated under Dependent Judgment conditions (Simulated Theft Experiment) and under Independent Judgment conditions. In the Dependent Judgment situation

TABLE I

PERCENTAGE OF CORRECT DECISIONS BASED ON EVALUATION OF TOTAL RECORD

DEPENDENT JUDGMENT (Simulated Theft Experiment)*			INDEPENDENT JUDGMENT (Present Experiment)**	
Judge	As Examiner	As Rater	Judge	As Rater
B	70	67	H	67
C	(33)	64	Y	71
D	77	--	Z	72
E	84	73	E	68
F	(100)	(100)		
Average	75	69		69

* Percentages in parantheses are based on fewer than 7 records; all others on 30 or more.

** Percentages based on 99 records.

most of the judges had two roles: as examiners they evaluated the records of suspects they themselves had tested; as raters they evaluated the records of other examiners. Thus, Judge E had an accuracy score of 84 when he made decisions on subjects he himself tested. His score dropped to 73 when he evaluated the records obtained by other examiners. Accuracy was further reduced to 68 when, more than a year later in the Independent Judgment Experiment, he reevaluated the same records. Judges H, Y, and Z served as raters only, since they were not involved in the Simulated Theft Experiment. Judge E was considered as a rater in the Independent Judgment situation: he did not know whose records were being used for this experiment, and he could not be expected to remember any details of the ratings or decisions he made more than a year ago.

It is apparent from Table 1 that there is no difference between the average accuracy of raters in the Dependent and Independent Judgment situations. The averages of the four raters in each experiment were identical, 69 percent. Assuming that the raters in both experiments were equivalent in overall ability, it may be concluded that the added information and the opportunity to compare records in the Dependent Judgment situation did not increase decision accuracy -- an unexpected conclusion. One explanation may be greater exposure to the records in the Independent Judgment situation. Each judge made four separate and independent evaluations of the records, first using the respiratory pattern alone, then the plethysmographic, then the psychogalvanic, and finally the total record with all its tracings. On the other hand, the judges in the Dependent Judgment situation arrived at their decisions after a careful examination and rating of the total record, but without intermediary decisions for each of the three physiological components. Although no time measurements were taken, it is safe to conclude that the decision time (per record) was shorter for the Dependent Judgment situation.

The "greater exposure" explanation, though seemingly reasonable, fails for Rater E who was involved in both experiments. In the Dependent Judgment Experiment his accuracy scores were 84 percent as examiner and 73 percent as rater. In rerating the same records one year later his accuracy score dropped to 68 percent, contrary to expectation. A possible explanation may be obtained from a

study of Rater E's decisions in both experiments. Of the 99 decisions in the Dependent Judgment situation, Rater E changed 29 of them under Independent Judgment conditions. This would seem to point to the existence of a large number of records (above 30%) which do not possess clear cut indications of diagnostic deception and which therefore do not "coerce" the same interpretation when reexamined after an appreciable time interval. With this explanation, emphasizing as it does a relatively large error variance, Rater E's poorer performance in the Dependent Judgment situation can be ascribed, in part, to a general regression effect. In addition, one may emphasize the loss of comparative clues to which Rater E may have become particularly sensitive in the Dependent Judgment experiment. Without these he became a more or less average rater in the Independent Judgment situation. He had been the best rater in the Simulated Theft Experiment.

Although Table I indicates that the examiners seem to be more accurate than either set of raters, the difference is not statistically significant. The table, however, suggests that amount of predecision knowledge available to raters may have an effect on variability of accuracy. A rough index for this conclusion may be found in the range of accuracy scores for each group. The group with most predecision knowledge -- the examiners who based their decisions on polygraph records, observations of suspects' behavior in the testing situation, comparison of three records -- had the largest range, 14 percentage points. The group intermediate in predecision knowledge -- the raters in the Dependent Judgment situation -- had the next largest range, 9 percentage points. The group with the least amount of predecision knowledge -- the raters in the Independent Judgment situation -- had the smallest range, 5 percentage points.

In summary, one definite conclusion is apparent. With sufficient time provided for evaluation (cf, exposure hypothesis) the accuracy of raters in the Independent Judgment situation is probably not much different from that of raters in a Dependent Judgment procedure.

CONFIDENCE IN DECISIONS

It was hypothesized that a lie detector operator in Independent Judgment situations would have less confidence in his decisions than if he worked under

Dependent Judgment conditions. In the latter case he would always have an opportunity to compare records of all suspects involved in a particular crime. Such comparisons were considered to generate more confidence in the resulting decisions than in others where this was not possible. Data from the present experiment were analyzed for possible evidence to test the hypothesis.

TABLE 2

AVERAGE CONFIDENCE RATINGS FOR CORRECT AND INCORRECT DECISIONS IN THE DEPENDENT AND INDEPENDENT JUDGMENT SITUATIONS

SCALE OF 0 - 6

ROLE	DEPENDENT JUDGMENT		INDEPENDENT JUDGMENT	
	Correct	Incorrect	Correct	Incorrect
T	3.85	3.07	3.78	3.61
L	3.62	3.35	3.87	3.79
I	3.94	3.40	4.17	3.03
Average	3.80	3.27	3.94	3.48

Table 2 presents the average confidence ratings in the two experimental situations. The confidence rating scale was the same as that used in the Simulated Theft Experiment. The results would seem to indicate that the hypothesis is not verified. On the average, the raters under Independent Judgment conditions gave higher ratings of confidence both for correct and incorrect decisions.

A further analysis was made of the confidence ratings of Rater E who was involved in both experiments. His confidence ratings for each record were compared with those of the other raters. Table 3 presents the results in terms of the percentage of times E's ratings were greater than, equal to, or less than the mean rating of his colleagues. The results were treated separately for the Dependent and Independent Judgment situations. It is apparent that E showed greater than average confidence in the decisions he made as a rater in the Dependent Judgment

TABLE 3

CONFIDENCE RATINGS OF RATER E (RELATIVE TO MEAN OF OTHER RATERS) IN THE DEPENDENT AND INDEPENDENT JUDGMENT SITUATIONS

COMPARISON (E vs. Mean Others)	JUDGMENT SITUATION DEPENDENT	INDEPENDENT
Greater	67	34
Equal	16	9
Smaller	17	57
	100%	100%

situation. A Chi-square test indicates that this is a statistically significant result (beyond the 0.01 level). Contrariwise, E manifested significantly lower than average confidence in the Independent Judgment situation. In fact, he was the most confident rater in the first situation, the least confident in the second.

As would be expected, E's confidence ratings dropped in absolute value from the first to the second experimental condition. In the Dependent Judgment situation the averages of his confidence ratings were 4.00 and 3.64 for correct and incorrect decisions respectively. The corresponding averages for the Independent Judgment situation were 3.72 and 3.41.

Why, then, would the other raters in the Independent Judgment situation have more confidence in their decisions than the raters in the Dependent Judgment situation? The most likely explanation concerns the notion of personal involvement. The raters in the Independent Judgment experiment were not personally involved in the records they were evaluating. They were not in the Simulated Theft Experiment; they did not know its weaknesses; they did not experience the wide range of response variability present in a highly motivated and emotionally charged experiment. On the other hand, the raters in the Dependent Judgment situation were personally involved in the conduct and execution of the Simulated Theft Experiment. It was their experiment, their subjects, their records. They knew the difficulties

involved and their rating attitudes were cautious and conservative. Because of this basic difference in attitude, there was a marked difference in the confidence they expressed in their ratings.

EXPERIENCE AND ACCURACY

As noted before two of the raters (E and H) in the Independent Judgment experiment were well trained both in polygraph testing and in interpreting polygraph charts. The other two raters (Y and Z) had no testing experience in actual lie detection experiments. They had, however, been trained to rate and interpret polygraph charts. But even in this, they had less experience than raters E and H.

TABLE 4

PERCENT ACCURACY SCORES OF RATERS (INDEPENDENT JUDGMENT SITUATION)

RATER	INDEX			
	Resp	Pleth	PGR	Total
E	45	59	69	68
H	47	65	68	67
Y	39*	39*	69	71
Z	35*	50	73	72

*Not significantly better than chance.

Table 4 presents the accuracy scores of the four raters for each of the physiological indices and for the total record. Thus, of the 99 records rated, E was correct in 45 percent of his decisions on the basis of the respiratory response alone. His accuracy increased to 59 percent when he based his decisions on the plethysmographic tracings. The highest accuracy was obtained with the psychogalvanic response (69%), better even than that for the total record where rater E had all three physiological tracings for evaluation.

The same pattern prevails for the entire table. Accuracy in detecting deception is least for the respiration pattern. The best accuracy is obtained with the

psychogalvanic response. Even when the total polygraph chart is examined, accuracy is slightly below that obtained for the psychogalvanic response alone.

As for the relation of accuracy and experience, the table shows that the more experienced raters (E and H) have higher scores for the respiration and plethysmographic indices. In fact, three of the four scores obtained by raters Y and Z on these indices are no better than chance. However, experience seems to have no influence on the accuracy with which the psychogalvanic response or the total record are evaluated. In fact, the less experienced raters have slightly better scores in these rating situations, but the difference is not statistically significant.

The results of this secion are not unexpected. Since the psychogalvanic tracing is less complicated than the plethysmographic and respiratory patterns, it lends itself to the development of more objective criteria in evaluating deception. Because of this, accuracy is no greater among more experienced raters than among less experienced, though well-trained, raters. Experience is of value in interpreting the more complicated respiratory and plethysmographic patterns as attested by the better accuracy scores of raters E and H. Finally, insofar as this experiment is concerned, use of the psychogalvanic response alone would have yielded results as accurate as those obtained from evaluating the entire polygraph chart with all three physiological tracings.

ERRORS OF MISCLASSIFICATION

Independent vs. Dependent Judgment. Three types of misclassification are possible: Thief and Lookout, Thief and Innocent, and Lookout and Innocent. In each, the misclassification is reversible, as for example, either mistaking the Thief for the Lookout (Thief-Lookout) or the Lookout for the Thief (Lookout-Thief). Table 5 presents the relative frequency of the six possible errors raters made in the Independent and Dependent Judgment situations. It may be observed that 11 percent of the errors in the Independent judgment situation were the mistakes of calling an Innocent Suspect a Lookout. This type of error comprised 15 percent of the total for the Dependent Judgment situation. The reverse misclassification (Lookout judged as Innocent) occurred in 16 percent of the errors in

the Independent Judgment experiment and in 15 per cent of the errors in the Dependent Judgment experiment.

TABLE 5

RELATIVE FREQUENCY (AS PERCENTAGES) OF MISCLASSIFICATION ERRORS FOR INDEPENDENT AND DEPENDENT JUDGMENT SITUATIONS

	DECISION (INCORRECT)					
ROLE	INNOCENT		LOOKOUT		THIEF	
	Ind	Dep	Ind	Dep	Ind	Dep
INNOCENT	--	--	11	15	13	5
LOOKOUT	16	15	--	--	26	34
THIEF	13	5	21	27	--	--

An overview of the table reveals that the most frequent errors were the Lookout-Thief misclassifications (26%, 34%) for both Independent and Dependent Judgment situations. Next in frequency were the Thief-Lookout errors (21%, 27%). In both misclassifications these errors were greater for the Dependent Judgment situation. The lowest frequency of misclassification occurred in the Innocent-Thief (5%) and Thief-Innocent (5%) decisions for the Dependent Judgment situation.

A relatively greater homogeneity of error is observed for the Independnt Judgment situation. The error percentage ranges from 11 to 26, a range half as great as that found among the Dependent Judgment percentages (5 to 34).

Probably the most critical result emerging from these comparisons is the relatively large number of Innocent-Thief and Thief-Innocent errors in the Independent Judgment situation. Furthermore, in this decision situation it is as easy to commit an Innocent-Thief error (13%) as an Innocent-Lookout error (11%), and almost as easy for the Thief-Innocent error (13%) as for a Lookout-Innocent error (16%). In contrast, the Thief-Innocent errors (5%) in the Dependent Judgment situation are much less frequent than the Lookout-Innocent or Innocent-Lookout errors (both 15%). The differentiation among the three roles seems to be an easier task in the Dependent Judgment experiment.

Among Physiological Indices. An informative comparison may be made of the six misclassification errors among the individual physiological indices. This will serve to point up the interaction of the various physiological indices with the six specific types of error. Table 6 presents the total frequencies of error found in the ratings of the three indices and in the ratings made on the total record, i.e., on the polygraph chart as a whole. Since there were no appreciable differences among the raters, the errors for each index were totalled and these sums comprise

TABLE 6

TOTAL FREQUENCIES OF EACH ERROR OF MISCLASSIFICATION FOR THE THREE PHYSIOLOGICAL INDICES AND FOR THE TOTAL RECORD IN THE INDEPENDENT JUDGMENT SITUATION

ROLE	INDEX		DECISION (INCORRECT)	
		INNOCENT	LOOKOUT	THIEF
	Resp	--	12	22
INNOCENT	Pleth	--	8	16
	PGR	--	13	16
	Total	--	13	16
	Resp	85	--	31
LOOKOUT	Pleth	55	--	34
	PGR	22	--	31
	Total	20	--	32
	Resp	76	11	--
THIEF	Pleth	65	11	--
	PGR	16	24	--
	Total	16	25	--

the data of the table. Thus, for the respiratory index there were 12 Innocent-Lookout errors while there were 85 of the Lookout-Innocent type.

The most striking feature of Table 6 is the magnitude of errors in the first column among the respiratory and plethysmographic indices. These errors involve the Lookout-Innocent and the Thief-Innocent misclassifications. These two mis-

classifications (of a total of six) account for 68 percent (161/237) of the total number of errors made with the respiratory index. The corresponding value is 63 percent (120/ 189) with the plethysmographic index. These errors are from three to four times as numerous as the corresponding errors involving the psychogalvanic response. In other words, when forced to use an index that yielded complex and vague criteria of deception (respiratory and plethysmographic), the rater would tend to judge a suspect as Innocent rather than incriminate him. And yet when a relatively more objective index (PGR) was introduced into the decision process, as can be observed in the "Total" Lookout-Innocent and Thief-Innocent errors, the misclassification was correspondingly reduced from 85 (Resp) to 20 (Total) and from 76 (Resp) to 16 (Total). A similar result is found for the Lookout-Innocent and Thief-Innocent errors with the plethysmographic index. The more easily rated and the more readily interpreted psychogalvanic index seems to have determined the final "Total" rating and thus dominated the decision process. The result was that the former Innocent ratings given on the basis of respiratory or plethysmographic tracings were now changed in the direction indicated by the psychogalvanic response.

INFLUENCE OF PGR ON RATINGS

One of the conclusions in the previous paragraph emphasizes the importance of the psychogalvanic response on the decisions of raters in their evaluation of the total polygraph chart. Table 4 indicates that the accuracy scores of raters using the phychogalvanic response alone do not differ more than two percentage points from the accuracy scores based on the total polygraph chart. Table 6 also indicates almost identical error frequencies for the psychogalvanic response and for the total polygraph chart. Table 7 presents the percentage of identical ratings (correct and incorrect) obtained by pairing the ratings made in each of the physiological indices with the ratings made on the total polygraph chart. Specifically, 97 percent of E's ratings based on the psychogalvanic response alone agreed with the ratings he made when he evaluated the total polygraph record. On the average, the percentage agreement between psychogalvanic reflex and total record ratings was 95 for the four raters. The average percentage of such agreement between

plethysmographic and total record ratings was only 58; that between respiratory

TABLE 7

PERCENTAGE OF IDENTICAL RATINGS WHEN TOTAL POLYGRAPH CHART DECISIONS ARE PAIRED WITH DECISIONS ON EACH PHYSIOLOGICAL INDEX

RATER	PAIRED DECISIONS		
	Total-PGR	Total-Pleth	Total-Resp
E	97	66	57
H	97	74	56
Y	93	39	49
Z	94	53	39
Average	95	58	50

and total record ratings still lower, 50. The more experienced raters (E and H) tended to get higher agreement scores for all three indices.

To conclude, the high degree of correspondence between accuracy scores for the psychogalvanic response and total record (Table 4) can be accounted for by the data in Table 7. Further evidence (Table 6) seems to indicate that the rating of the total polygraph record was relatively uninfluenced by the respiratory and plethysmographic evidence that may have been present in the chart. Reliance was placed almost entirely on the psychogalvanic index which influenced the final decision.

CONCLUSIONS

1. Decision accuracy in the Dependent Judgment situation was no greater than that attained under Independent Judgment conditions. Greater exposure to the records in the Independent Judgment situation probably counterbalanced the inherent advantages assumed to be present in the Dependent Judgment case.
2. The hypotheses that confidence in decisions would be consistently greater for the Dependent Judgment situation was not verified for the group data.
3. In the case of the one rater who served in both experiments, accuracy and

confidence in decisions decreased from the Dependent to the Independent Judgment situation.

4. Experienced raters were more accurate than the less experienced raters in analyzing respiratory and plethysmographic indices for evidence of deception. No difference in accuracy between the two groups of raters was noted in the evaluation of the psychogalvanic response or of the total polygraph chart.
5. The more "serious" errors of misclassification (Thief-Innocent and Innocent-Thief) were more frequent in the Independent Judgment situation.
6. In using the less objective indices (respiratory and plethysmographic), raters tended to judge Thief and Lookout as Innocent approximately 3-4 times more frequently than with the psychogalvanic index.
7. The psychogalvanic response determined the final decision in the analysis of the total polygraph chart. Furthermore, greatest accuracy was attained when the psychogalvanic response alone was used in the lie detection decision.

SECTION II

ACCURACY OF MEASURED AND RATED PHYSIOLOGICAL RESPONSE SYSTEMS USED IN LIE DETECTION WORK

The decisions made by lie detector operators are basically subjective in character. Undoubtedly they are based on careful study of the polygraph charts but usually there are no measurements, no statistical analyses, and no specific objective criteria against which the measurements are compared.

In the previous section, lie detector operators rated the "significance" of the physiological reactions to each of the critical questions used in the interrogation. This was done independently for each index: respiratory, plethysmographic, and psychogalvanic. When this analysis was completed, the lie detector operator was instructed to give his overall decision as to the guilt, complicity, or innocence of the individual whose records he had just rated. Despite this attempt to provide a firm basis for his final decision, the process was essentially subjective since only a visual comparison of the tracings was required. There were no measurements made of the physiological responses.

If computer techniques were to be utilized, the visual evaluation would have to be superseded by objective measurement. The measurements would have to be based on those asp ts of the visual record which provide the operator with the subjective criteria he uses in arriving at his judgment. Once such measurements were made, they could be used with complete objectivity to determine the guilt, complicity, or innocence of the individual tested. The accuracy thus attained could be compared with that achieved by the lie detector operators evaluating the same records. If the accuracy of the objective measurements were comparable to that of the lie detector operators, computerization would be feasible. With the physiological signals converted to digital form, the examination of a suspect could be facilitated by "immediate" feedback from the computer indicating the minute-to-minute (or the cumulative) status of the suspect's total physiological reactivity.

STATEMENT OF PROBLEM

Since polygraph records were available from the previous study (Kubis 1962), these could be subjected to measurement. The first problem was to determine the

most feasible and reliable characteristics of the physiological reactions. Once these were measured and combined into a diagnostic form which would provide a decision as to the guilt, complicity, or innocence of a suspect, the final and basic question could be answered: Will objective measurements provide the same degree of decision accuracy as lie detector operators?

If the decisions of lie detector operators were found to be more accurate than those derived from purely objective measurement, more work would have to be done either on objectifying the subjective criteria or on discovering other measureable physiological characteristics that would increase the accuracy of the objective decisions.

PROCEDURE

There were three phases to the procedure: the characteristics to be measured had to be selected: a sample of records had to be obtained; the method of evaluating the accuracy of the objective (measurement) and subjective (lie detector operator ratings) methods had to be determined.

MEASURED CHARACTERISTICS

The three physiological reactions -- psychogalvanic, plethysmographic, and respiratory -- differ greatly in form and complexity. The description of the characteristics selected for study is presented in separate sections for each reaction. A detailed analysis of the measurement procedure is included in Appendix A.

Psychogalvanic Reaction. Two measurements were used to serve as indices for the psychogalvanic reactions. These were the height of the response and its "width." The height of the deflection is a function of the conductance. "Width" measures recovery time. Since it was not always possible during the testing period to have the psychogalvanic deflection return to its base line, recovery time was measured at that point of the curve where the return sweep of the deflection was one-half the maximum height attained. This criterion made it possible to get a measure on all the deflections used in the study.

Plethysmographic Reaction. Two of the characteristics of the plethysmographic reaction are direct analogues of the height and width mentioned above. In excite-

ment the change in finger blood volume is indicated by a rise in the plethysmographic curve. Within a short period of time the curve returns to its base line. Consequently, amplitude or height can be measured; similarly, recovery time or width. In addition, the change in the magnitude of the pulse beat was also used. To facilitate later discussion, these three characteristics are referred to as Height, Width, and Change.

Respiratory Reaction. It was felt that the amplitude and frequency of the respiratory cycles contained all the relevant information that would reflect the emotional state of the subject under test.

The selection of these seven characteristics was based on the diagnostic significance they were considered to possess. In particular, the height of either the psychogalvanic or plethysmographic reactions has always been considered a good indicator of the "disturbed" or emotional state of the individual at that point. Both are used by lie detector operators as presumed indices of disturbance (or lying, if properly interpreted). Similarly a diminution of respiratory amplitude at a critical question has often been used as an index of lying. Other characteristics of the physiological reactions were not selected for analysis because they failed to meet the criteria of measurability and diagnostic significance for detecting deception.

Change in responsivity is the critical index for Joseph (1957). The most obvious measure of change is a comparison of the reaction at a critical point with the reactions before and after it. Such was the procedure used. As an example, the Height of the psychogalvanic reaction to a critical question was divided by the sum of the Heights to the noncritical questions before and after it. (Averaging the Heights of the two noncritical questions would have introduced a constant factor of 0.5, common to all measurements and therefore an unnecessary operation.)

All measurements were done by two statistical clerks who did not know the nature or purpose of the experiment. There was a preliminary training period to assess the adequacy of the measurement instructions and to develop consistency and reliability in the measurement procedure.

THE SAMPLE OF RECORDS

The measurement of the seven characteristics was very time-consuming. Con-

sequently, only a limited sample was selected to serve as a pilot indicator of the diagnostic promise inherent in the objective measurements. The records used for the objective analysis were chosen from the second half of the Simulated Theft Experiment (Kubis, 1962). They comprised 11 complete experimental groups of three persons. Each such group contained a Thief, a Lookout, and an Innocent Suspect. All of these groups (totalling 33 persons) had been examined by one lie detector operator thus insuring relative uniformity of questioning and machine operation. These records had been analyzed and rated by three persons: the examiner and two raters.

ACCURACY EVALUATION

Lie Detector Ratings. The physiological reactions to each critical question (i.e., a question relating directly to the Simulated Theft) were rated on a scale of 0-3 to indicate the degree of disturbance the question aroused. The critical response (reaction to the critical question) was compared with its predecessor and with its successor. Depending on the comparative magnitude of the disturbance aroused by the question, the critical response was given one of the following numerical ratings:

3 - very significant
2 - significant
1 - doubtfully significant
0 - nonsignificant

This scale was used and described in the Simulated Theft Experiment (Kubis, 1962). These ratings were combined into three discriminant scores: the Thief-Innocent (T-I), the Thief-Lookout (T-L), and the Lookout-Innocent (L-I). These scores were to determine the relative accuracy of the three types of discriminations possible within a group of three persons one being a Thief, one a Lookout, and one an Innocent Suspect. Thus, for example, the T-I score was constructed so as to distinguish the Thief from the Innocent Suspect. With three physiological reactions, there were three T-I scores, one for each of the indices: the respiratory, the pletnysmographic, and the psychogalvanic. In the earlier research (Kubis, 1962) it was found that the most accurate discriminator was the psychogalvanic response. The least accurate was the respiratory response.

The natural question that arises is, Would a combination of the three physio-

iogical indices increase accuracy? The simplest type of combination, the sum of the three physiological discriminants, proved no more accurate than the single psychogalvanic discriminant. However, the use of linear discriminant function analysis provided a set of weights (or multipliers) for the physiological discriminants that maximized the efficiency of classification. This linear function proved to be the most accurate discriminant.

For any required discrimination, as, for example, the classification of an individual as a Thief or as an Innocent Suspect (T-I), there were five sets of discriminant scores: one for each of the physiological responses, one for the sum of the three physiological discriminants, and finally the maximizing linear discriminant function. This was the case also for the T-L and for the L-I scores.

Decisions Based on Measurements. Although the same three discriminations (T-I, T-L, and L-I) must be made whether the physiological curves are rated or measured, there are a number of differences that must be mentioned. In the one case the physiological tracings are evaluated and rated by eye; in the other, the same tracings are measured on a scale. In the subjective evaluation, the total physiological pattern (ex. respiration) accompanying a question is compared with the total physiological patterns (ex., respiration) accompanying the surrounding questions. In the objective procedure only two facets of the particular curve (ex., amplitude and frequency of respiration) are singled out for measurement. Although it appears that there is potentially more information in the subjective evaluation, it must be admitted that the measured information is more reliable. Finally, the multiple measurements made on each physiological response make possible many different linear combinations of measurements. Specifically, there are 12 different (3x2x2) linear discriminant scores that have exactly one measurement from each physiological reaction. Further, theoretically there is no inherent restriction on the number of variables to combine. There may be as few as two or as many as seven. In the present case the emphasis has been on linear combinations utilizing one measurement from each of the physiological reactions. Some additional linear discriminants were computed and these will be indicated in the treatment of results.

Accuracy scores, for both the rated and measured conditions, will be expressed

in terms of percent correct discriminations. The discriminations will be Thief vs. Innocent (T-I), Thief vs. Lookout (T-L), and Lookout vs. Innocent (L-I). In this way, it will be possible to evaluate the relative accuracies of the three types of decisions that are inherent in the indentification of three members of a group one of whom is a Thief, one a Lookout, and one an Innocent Suspect.

RESULTS

The basic variables under study were the three physiological reactions to "critical" questions used in the Simulated Theft Experiment. The reactions to these questions were evaluated in two ways: by direct physical measurement of the tracings with respect to such characteristics as Height, Width, Change, and by a visual examination of the same tracings by trained lie detector operators who rated the significance of the reactions on a scale of 0-3. Objective measurement analysis yielded at least two indices for each physiological reaction, e.g., Height and Width for the psychogalvanic response, Frequency and Amplitude for respiration, and Height, Width, and Change for the plethysmographic tracing. The visual analysis by lie detector operators produced one overall rating for each of the physiological response systems.

Since the measurements and the ratings were obtained from the same set of 33 polygraph charts, a direct comparison of the accuracy of the two methods (measurement vs. rating) was possible. Accuracy was expressed in terms of percent: the percent of correct discriminations between pairs of subjects one of whom was a Thief, the other an Innocent Suspect (the T-I discrimination); the percent of correct discriminations between Thief and Lookout (the T-L discrimination); and the percent of correct discriminations between Lookout and Innocent Suspect (the L-I discrimination).

In the sections that follow, the initial comparisons between the measured and rated data will focus on the accuracy of the single physiological indices. The subsequent comparisons between the measurement and rating procedures will involve the accuracy scores attained by combining indices.

SINGLE PHYSIOLOGICAL INDICES

The first comparison between the two methods of scoring, objective measurement and visual rating, involves the accuracy attained by using single indices. Table 8 presents the accuracy scores of the measurements and ratings for each of the physiological reactions. Each measured percentage is based on 11 paired discriminations. In other words, the 91 percent accuracy attained by using measured Height of the psychogalvanic response to make the T-I discriminations indicates that in 10 of 11 comparisons the psychogalvanic index was larger for the Thief than

TABLE 8

ACCURACY SCORES FOR SINGLE PHYSIOLOGICAL INDICES OBTAINED BY MEASUREMENT AND BY RATINGS

MEASURED AND RATED INDICES	DISCRIMINATION T-I	T-L	L-I	GENERAL AVERAGE
PSYCHOGALVANIC				
Measured				
Height	91	91	100	94
Width	82	91	82	85
Visual Rating	91	90	82	88
PLETHYSMOGRAPHIC				
Measured				
Height	64	55	82	67
Width	55	55	82	64
Change	82	73	64	73
Visual Rating	82	77	73	77
RESPIRATORY				
Measured				
Frequency	64	45	45	52
Amplitude	55	55	64	58
Visual Rating	71	41	71	61

for the Innocent Suspect. The accuracy of visual ratings for the same 11 Thief-Innocent pairs is expressed as 91 percent and indicates that in 30 of 33 comparisons of Thief-Innocent pairs the psychogalvanic rating was greater for the Thief than for the Innocent Suspect. There were 33 comparisons in the rating because three lie detector operators rated the polygraph charts of the 11 Thief-Innocent pairs. For all visual ratings, then, the percentages are based on the evaluations of three raters.

The overall picture indicates that greatest accuracy is attained for the psychogalvanic response, whether it be for the measured data or for the rated data. Least accurate are the respiratory indices, measured or rated. Approximately midway lie the accuracy scores for the plethysmographic response.

The main purpose of measuring the physiological reactions was to determine how accurate discriminations could be when certain selected aspects of the total reaction pattern were used as diagnostic indices. Such accuracy was to be compared with the accuracy of ratings of lie detector operators who evaluated the total reaction on the basis of a visual examination of the curves. Thus, as regards the Thief-Innocent discrimination the measured Height of the psychogalvanic response proved to be as accurate (91%) as the ratings of the lie detector operators who studied the total psychogalvanic pattern in arriving at their rating of the same response. Measured Width (82%), however, did not prove to be as accurate as the Visual Rating (91%). It is likely that the lie detector operators are more influenced in their ratings by the height of the psychogalvanic response rather than be its width (recovery time). "Insofar as the psychogalvanic response is concerned, when all three types of discrimination are averaged, the measured height yields the greatest accuracy (94%)". Visual ratings (88%) are slightly more accurate on the average than measured Width (85%). The important fact that emerges from this analysis is that measured Height alone is at least as accurate as the Visual Rating, despite the greater amount of information potentially available in the visual evaluation of the total physiological pattern.

A study of plethysmographic accuracy reveals that the average of Visual Ratings (77%) is slightly higher than the average of Change in pulse beat (73%). Height (67%) and Width (64%) of plethysmographic response are, in turn, slightly less

accurage then Change. The pertinent observation is that only one measured aspect of the plethysmographic pattern (Change) is almost as accurate as the Visual Rating which is based on the total plethysmographic reaction.

A similar result is to be noted for the respiratory response system which attained the lowest degree of discriminatory accuracy. Measured Amplitude had an average accuracy of 58%, a value just slightly lower than the 61 percent for Visual Rating.

In summary, there is at least one measured characteristic in each of the physiological response systems that attains an accuracy score very close to that achieved by the visual ratings of lie detector operators. It is thus within the realm of practicality to replace such subjective ratings by objective measurement without sacrificing overall accuracy. Further, since the terminal decisions of lie detector operators are not significantly more accurate than the optimal weighting system assigned to their ratings of individual physiological reactions, it is theoretically conceivable that the objectively measured responses -- ultimately done under computer control -- can be optimally weighted by a computer into an objective decision reflecting the guilt or innocence of a subject.

COMBINATION OF SCORES

It was noted above that the measurement procedures yielded two scores for the psychogalvanic response, three for the plethysmographic response, and two for the respiratory response. There were, then, twelve possible ways of obtaining a combined score by always selecting one score from each of the three physiological response systems. As an example, psychogalvanic Height, plethysmographic Change, and respiratory Amplitude could be used to determine the degree of accuracy such a combination would have in discriminating between a Thief and an Innocent Suspect (T-I), between a Thief and a Lookout (T-L), and between a Lookout and an Innocent Suspect (L-I). Two ways were used to combine such scores: simple summing of the individual scores or weighting each score by means of a linear discriminant function. These two will be called Summed Score and Discriminant Score. The linear discriminant procedure was used and described in

the Simulated Theft Experiment (Kubis, 1962).

There was only one rating for each of the physiological indices. It was based, as mentioned earlier, on an overall evaluation of the total pattern involved in each physiological response. With only one rating available for each physiological response, only one combination of all three was possible. The two methods of weighting such a combination were the same as indicated above: Summed Score and Discriminant Score. In this case it was the ratings that were summed or weighted by a linear discriminant function.

It would serve no useful purpose to catalogue all 24 measurement scores (12 Summed, 12 Discriminant), each a combination of the three physiological parameters. The accuracies with which these combined scores were able to make the T-I, T-L, and L-I discriminations have been averaged and the results presented together with the two combined Visual Rating scores (one Summed and one Discriminant) in Table 9. The overall results are fairly clear. The scores obtained by measurement, when combined so as to include one representative from each of the physiological reactions, yield accuracy scores that are slightly better on the average than the combined visual ratings obtained from the lie detector operators. Thus, when simply summed, the measurement scores attain an average accuracy of 87 percent, two units higher than the corresponding summed ratings (85%). The dis-

TABLE 9

PERCENT ACCURACY OF THE COMBINED MEASUREMENT SCORES AND THE COMBINED VISUAL RATINGS FOR THE THREE TYPES OF DISCRIMINATION

DISCRIMINATION	VISUAL RATINGS		MEASUREMENT SCORES	
	Summed	Discriminant	Summed	Discriminant
THIEF-INNOCENT	91	94	90	83
THIEF-LOOKOUT	88	94	82	92
LOOKOUT-INNOCENT	76	79	89	97
General Average	85	89	87	91

criminant weighted scores (91%, 89%) are slightly and uniformly better in accuracy than the summed scores for both the measurements (87%) and ratings (85%). The superiority of the averaged measurement scores is due in large part to the differential accuracy noted for the Lookout-Innocent discrimination in which the Visual Rating accuracy happened to be relatively poor.

This analysis is intended to be suggestive rather than exhaustive. The percentages are based on only 11 paired comparisons within each of the three types of discrimination. Despite this limitation, the results are encouraging from at least two points of view. In the first place objective measurement yields results that can be used to discriminate among Thief, Lookout and Innocent Suspect with at least the accuracy obtained from ratings of lie detector operators. The accuracy percents for the various discriminations range from 82 to 97 for the combined measurements. It is apparent that the measurements are tapping real physiological differences in the responses of the various groups who had different roles to play in the Simulated Theft Experiment.

It may also prove instructive to combine the several measurements within each physiological response to discover how accuracy is affected by including more than one measurement aspect in the discrimination task. With this objective the two scores for the psychogalvanic response, Height and Width, were combined by simple summing and by weighting the two scores with a linear discriminant function. This was also done for the three scores (Height, Width, and Change) obtained from the plethysmographic response and for the two scores (Amplitude and Frequency) from the respiratory reaction. The accuracy in discrimination (T-I, T-L, L-I) for each physiological combination is presented in Table 10. A comparison of these results with those of Table 8 does not reveal any consistent increase in accuracy of the combined scores over that found for the single scores. Thus, one would do as well with PGR Height alone as with a combination of Height and Width. For the plethysmograph, however, the discriminant scores in the T-I and the T-L discriminations would do better than either of the three single scores. But this is not true for the L-I discrimination. As for respiration, only in the T-I discrimination is there any appreciable increase in accuracy for the combined scores. The ab-

TABLE 10

PERCENT ACCURACY FOR THE COMBINATIONS OF SCOR THIN EACH PHYSIOLOGICAL RESPONSE SYSTEM FOR THE THREE KINDS O DISCRIMINATION

COMBINATION	THIEF-INNOCENT		THIEF-LOOKOUT		LOOKOUT-INNOCENT	
	Summed	Discriminant	Summed	Discriminant	Summed	Discriminant
PGR (Height, Width)	91	82	100	91	91	82
PLETHYSMOGRAPH (Height, Width, & Change)	73	100	64	91	91	73
RESPIRATION (Frequency, & Amplitude)	55	73	55	55	55	45
	73	85	73	79	79	67

sence of appreciable increases in accuracy for the combinations is due in part to the relatively high degree of correlation between the indices within the physiological response systems.

CONCLUSIONS

1. Measured characteristics of physiological responses can attain an average accuracy equivalent to that achieved by visual ratings obtained from lie detector operators. In other words, there is at least one aspect of a physiological response, e.g., height of PGR tracing, that can be used to discriminate between a Thief and an Innocent Suspect with the same degree of accuracy as that achieved by ratings of lie detector operators who examine the total psychogalvanic response pattern in arriving at their evaluations. This is generally true of the plethysmographic and respiratory responses as well.
2. The combinations of the measured indices within each physiological response system, e.g., intensity (Height) and recovery time (Width) of the psychogalvanic tracing, do not yield appreciable and consistent increases in accuracy over those attained by the single indices.
3. The combinations of the measured indices, one from each of the three physiological response systems, yield an average accuracy of discrimination at least

as large as that attained by the corresponding combination of rated physiological reactions.

Although these results must be evaluated against the background of limited sample size, it is encouraging to note that the ratings of lie detector operators are not more diagnostic than the objective measurements that are most likely possible with the aid of a computer. More work needs to be done on the nature and frequency of "serious" errors (e.g., calling - Innocent Suspect a Thief) in the objective measurement system.

REFERENCES

Joseph, C. N. Analysis of compensatory responses and irregularities in polygraph chart interpretation. In V. A. Leonard (Ed.) Academy Lectures on Lie Detection. Vol. I. Springfield, Ill.: Thomas, 1957, Pp. 93-99

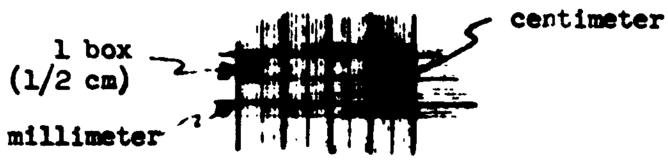
Kubis, J. F. Studies in lie detection: Computer feasibility considerations, U. S. Air Force, RADC-TR 62-205, 1962.

APPENDIX

DIRECTIONS FOR OBJECTIVE MEASUREMENT OF RESPONSES

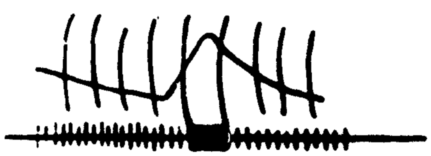
GENERAL INSTRUCTIONS FOR OBJECTIVE MEASUREMENT OF RESPONSES

1) Use the glass grid provided to make all measurements which cannot be made directly from the lines marked on the record paper. This grid is ruled in millimeters, half-centimeters, and centimeters, as shown in the diagram. The half-centimeter square will hereafter be referred to as a "box."

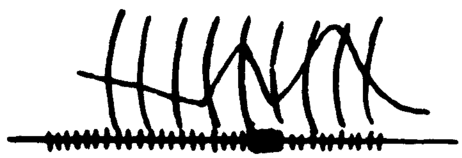


2) Each question is identified by a solid block on the bottom line of the record, as shown. The responses, starting immediately above these blocks are the ones to be measured.

A response to a question is considered valid, even if the response slightly precedes the solid block on the record. If, however, the response occurs a full box (1/2 cm.) or more before the block, measure the next response.

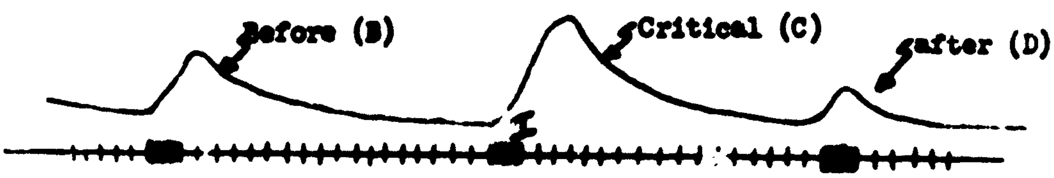


Count this as the response



In this case, count the second rise as the response.

3) Measure only the response marked by the Roman numeral (critical question), and the response immediately before and after this question. Record the values in the appropriate columns marked on the data sheets, either column B (before critical), column C (critical), or column D (after critical).



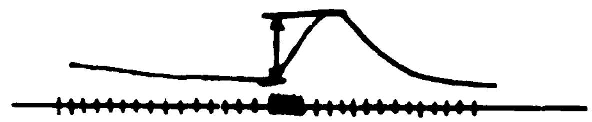
4) Make all measurements to the nearest 1/2 millimeter.

5) Be sure to note the order of questions on the record sheets: some are ordered I, II, III; other III, II, I; other II, I, III; etc., and record in the appropriate place on the data sheet.

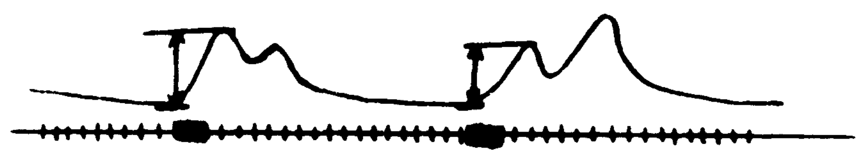
SPECIFIC INSTRUCTIONS: CRITERIA FOR MEASUREMENT

PGR Height

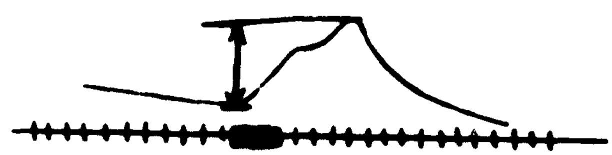
Measure from beginning of rise to top of initial rise.



When there is a double response, measure only the first one, even if the second one is higher.

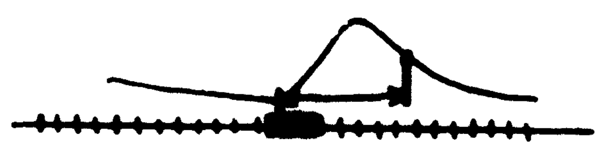


If the first bulge does not show definite signs of moving down, include the second one in the measurement.

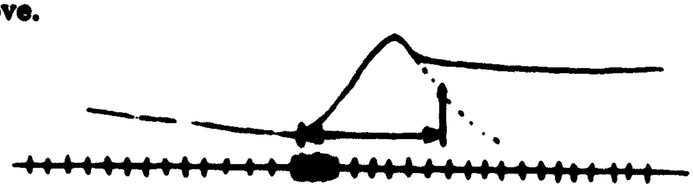


PGR Width

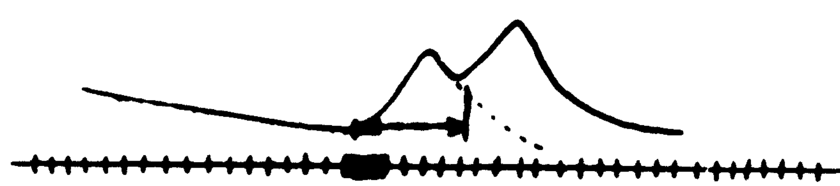
Measure the horizontal distance from the beginning of the rise to the point where the curve has fallen one half the height of the rise.



When the curve does not fall to the half-way point, extrapolate it and measure as described above.



When there is a double response, extrapolate the first (if necessary), and measure it as above.



Plethysmograph Frequency

Count the number of spikes per 5 boxes (25mm.). Always count at the bottom of the spikes as shown in the example. In this example, there are 14 spikes in the 5 boxes.

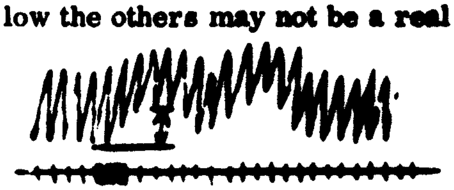
Be sure to include as many spikes as possible after the question, by placing, the first box exactly on the point of the first spike, as shown in the example above, otherwise you may miss a spike or two in your count.

If the next question occurs before 5 boxes have elapsed, use as many boxes as possible in your measurement, but keep the number of boxes used, constant for each B-C-D triad.

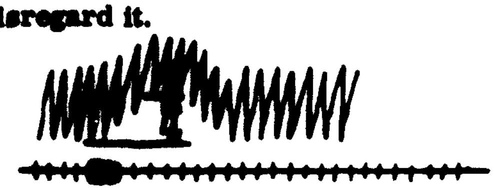
Plethysmograph Height Of Rise

Measure the height of the rise from the two beginning points (prior to the rise) to the two shallow points of the rise. If the level of the two points does not coincide, estimate their mean and measure this distance.

In the case of a double rise, measure only the first one. Unless the rise shows definite signs of dropping, consider it as a single rise, i. e., a single spike below the others may not be a real drop, so disregard it.

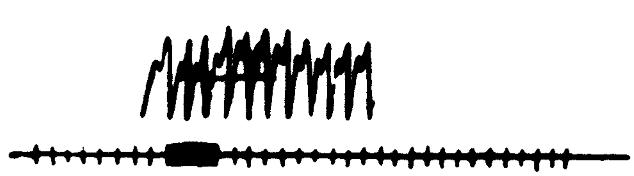


This is a double rise



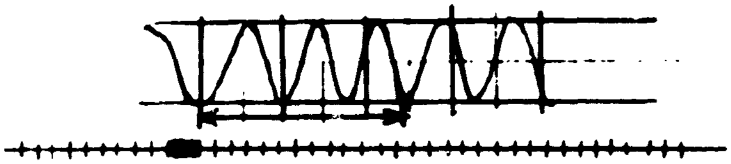
This is not a double rise

If no rise is evident, check for a notch in the middle of the spike and measure the rise in these notches, if any.



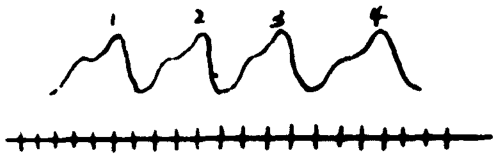
Respiration Frequency

Count the number of millimeters on the grid within three cycles. If the limits of any one question in a triad include only two (or less) cycles, then count the number of millimeters for that number of cycles, but keep it constant for all three questions in each triad. Never use parts of cycles.

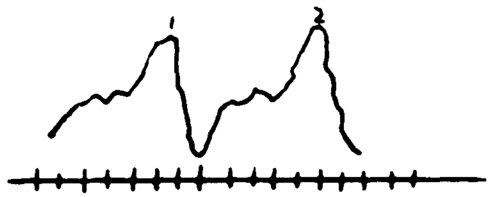


There are 24 millimeters in this example

Count only clear, evident cycles



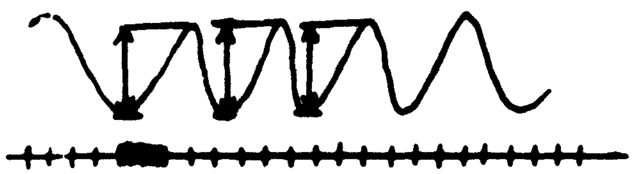
This example has four cycles



This one has only two

Respiration Amplitude

Add the heights of all three cycles in each question of a triad. If there are less than three cycles before the next response, use as many as possible but keep the number of cycles used constant for each triad.

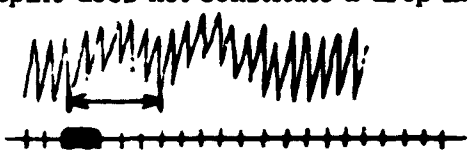


Measure the left side of the cycle in all cases.

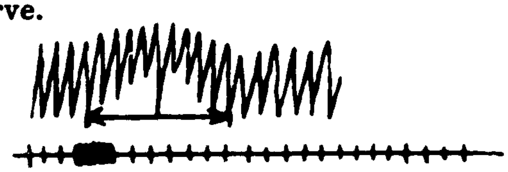
Plethysmograph Width Of Rise

Measure the horizontal distance from the beginning of the response to the end of the response. To avoid chance results, always make sure there are at least two low points at both the beginning and end of the rise.

In case of a double response measure only the first one. Remember, a single spike does not constitute a drop in the curve.



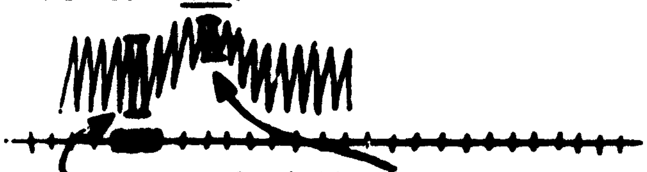
This is a double response



This is not a double response

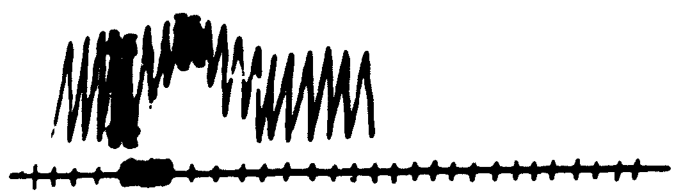
Plethysmograph Change In Pattern

Divide the length of the first two responses by the length of the two shortest successive responses for each question in the triad. That is, measure the height of the spikes; add the heights of the first two and divide by the sum of the heights of the shortest two.



These two divided by these two

In making the measurements of height, measure the height of the right side of the spike.



analysis and measurement of the three physiological response systems. The acteristics of the psychogalvanic response selected for measurement were r change in resistance (Height) and recovery time (Width). Amplitude and freq were the measured indices obtained from the respiratory tracings. Height, Widt Change were measured from the plethysmographic response. The accuracies these indices, separately or in combination, were compared with the accura attained by the ratings of lie detector operators who evaluated the total respo pattern of each physiological response in arriving at their ratings.

The measured characteristics of the physiological response systems were to be as accurate as the ratings of the lie detector operators in discriminatin tween culprit, collaborator, and innocent suspect. Continued research should it possible to objectify most of the lie detection indices with the aid of a com

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Dr. J. F. Kubis Fordham University New York, New York	2a. REPORT SECURITY CLASSIFICATION 2b. GROUP

3. REPORT TITLE

QUANTITATIVE ANALYSIS OF POLYGRAPHIC DATA

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report March 1962 - July 1964

5. AUTHOR(S) (Last name, first name, initial)

Kubis, Joseph F.

6. REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
January 1965	35	2

8a. CONTRACT OR GRANT NO. AF30(602)-2634	9a. ORIGINATOR'S REPORT NUMBER(S) None
b. PROJECT NO. 5534	
c. Task 553401	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d	RADC-TDR-64-101

10. AVAILABILITY/LIMITATION NOTICES

Report released to Office of Technical Services for sale to public.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Rome Air Development Center QRC, Applied Science Section Griffiss AFB NY 13442

13. ABSTRACT Two types of decision situations are characteristic of lie detection investigations: the dependent judgment case in which the examiner, after comparing all records, selects the guilty individual (and possible accomplices) from among a group of suspects known to include the culprit(s); and, the independent judgment case, in which a decision of innocence or guilt is made independently for each suspect on the basis of his record alone. In the latter situation the suspects are usually apprehended one at a time and at irregular intervals.

Rater accuracy for each decision situation was evaluated by utilizing 100 records obtained in the Simulated Theft Experiment (Kubis, 1962), a dependent judgment situation. These records were evaluated under independent judgment conditions. It was anticipated that the opportunity of comparing the records of all suspects in the dependent judgment situation would result in greater accuracy than that attainable in the independent judgment situation.

The results indicate that neither accuracy of decisions nor confidence in them was diminished under independent judgment conditions. The more "serious" errors of misclassification were more numerous in the independent judgment situation. Greatest accuracy was achieved with the psychogalvanic index of deception, and this index tended to determine the direction of the final decision in the analysis of the total polygraph chart.

Records of 33 subjects from the Simulated Theft Experiment were selected for further

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Physiology						
Emotion						
Stress						
Psychogalvanic						
Pulse						
Respiration						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (*TS*), (*S*), (*C*), or (*U*).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

UNCLASSIFIED
Security Classification